DOCUMENT RESUME

ED 161 675 SE 024 989

AUTHOR	Welch, Wayne W.; Gullickson, Arlen R.
TITLE	A Strategy for Evaluating the NSF Comprehensive Program for Teacher Education. Research Paper No. 1.
INSTITUTION	Minnesota Univ., Minneapolis. Coll. of Education.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	[75]
GRANT	NSF-GW-6800
NOTE	16p.; For related documents, see SE 024 990-999 and ED 148 632-640; Contains occasional light and broken type
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Design; Educational Research; *Elementary Secondary Education; Higher Education; *Inservice Teacher Education; Mathematics Teachers; *Program Evaluation; *Science Education; Science Teachers; *Summative Evaluation
IDENTIFIERS	*Minnesota Research & Evaluation Project; National Science Foundation; Research Reports

ABSTRACT

This document presents the design of an evaluation study for the Program of Comprehensive Grants for Teacher Education which is funded by the National Science Foundation (NSF). The main concerns of the evaluation were the need to provide context information to individual project directors and to develop a summative strategy for the national program. A total of 17 instruments were selected to collect a variety of cognitive and affective data for both science and mathematics projects from students, teachers and administrators. Approximately one-half of these instruments were developed for this evaluation while the remainder were chosen from existing instruments. A total of 354 schools from five experimental regions and five control regions participated in this study. The data gathered by this national design will be processed and analyzed during the next several months and presented to Comprehensive Project Directors and to the NSF. (HM)

RESEARCH PAPER #1

A Strategy for Evaluating the NSF Comprehensive Program for Teacher Education

Wayne W. Welch and Arlen R. Gullickson
University of Minnesota

This study was supported by grant GW-6800 from the National Science Foundation to the University of Minnesota. Wayne W. Welch, Project Director.

A STRATEGY FOR EVALUATING THE NSF COMPREHENSIVE PROGRAM FOR TEACHER EDUCATION

Wayne W. Welch and Arlen R. Gullickson*

University of Minnesota

Among the more innovative of recent NSF program changes is a venture entitled the Program of Comprehensive Grants for Teacher Education. The shortened label for this program is "Comprehensives," Although the program has been in existence less than two years, it is gaining the status and attention of other well-known NSF teacher education programs, such as the Summer and Academic Year Institutes, and the Cooperative College-School Science (CCSS) program.

The goals of the Comprehensive Program are twofold:

1. To help schools, through the education of their instructional, resource and supervisory personnel, in developing their capacity for self-improvement in science and mathematics education, and

2. To assist the efforts of colleges and universities in developing as part of their regular activities more effective programs for the pre-service and in-service education of science and mathematics teachers.

Five experimental projects were funded under this program in 1972. The projects were funded for four years. Mathematics projects were established at Notre Dame and San Jose. Science Comprehensives were located in Mississippi, South Dakota, and Wyoming. These five projects are the focus for the evaluation strategy described in this paper.

The unique features of the Program that should be mentioned here include: (1) a regional rather than national focus, (2) active responsiveness to school needs, (3) program flexibility, and (4) multi-year funding. A fifth distinguishing characteristic of the Comprehensive Program is its attention to ongoing planning and evaluation. It is this last point that serves as the substance of this paper. What strategy would be appropriate for evaluating the impact and efficacy of the Comprehensive Program?

*The authors are indebted to Alphonse Buccino, Glenn Bracht, James Rutherford, and Herbert Walberg for comments on an earlier draft of this article.

The underlying consideration in conducting this, or any other, evaluation is that its function is to gather information for the purpose of making decisions (Welch, 1969). Information pertaining to two primary decision situations was sought in the present study: (1) regional needs assessment data to be used by project directors in developing the various components of their comprehensive project, and (2) pretest data against which potential posttest changes could be detected.

The main concerns of the evaluation were the need to provide context information to individual project directors and to develop a summative strategy for the national program. An evaluation strategy to address these problems is described below.

Evaluation Plan

To obtain data bearing on the several objectives of the program, a multi-level testing program was implemented. A variety of cognitive and affective data was sought for both science and mathematics projects from students, teachers, and administrators. The final selection of instruments was based upon the underlying philosophy of the Comprehensive Program (Rutherford, 1971), availability of instruments, and advice of an external Advisory Committee. (The latter group consisted of six members selected from the areas of science and mathematics education, educational psychology, school systems, and the NSF.)

From more that three dozen instruments considered, a total of 17 were selected. Approximately one-half were developed for this evaluation while the remainder were chosen from existing instruments. These instruments are listed in Table 1 together with the appropriate region.

Insert Table 1 about here

A brief description of each instrument follows.

TABLE 1

Evaluation Instruments

	Principals	Teachers	Students
cs Notre Jose onding egions)	Principal Questionnaire Work Values Inventory Mathematics Inventory	National Teachers Exam-Mathematics* Mathematics Inventory for Teachers Teacher Questionnaire	Mathematics Achievement Test Mathematics Attitude Inventor Learning Environment Inventor
Wyoming, pi, South d their egions)	Principal Questionnaire Science Process Inventory Science Attitude Inventory	National Teachers Exam-Science* Science Process Inventory Science Attitude Inventory Teacher Questionnaire	Test of Achievement in Scienc Science Process Inventory Science Attitude Inventory Learning Environment Inventor

ven to control sample

1. Test of Achievement in Science (NAEP, 1970)

   A. Form I
   A 45-item multiple choice test for secondary school science comprised of items selected from the National Assessment Test for Science. Items were selected with difficulty levels between 5% and 90% from those released in 1971 for 17 year-olds.

   B. Form II
   A 40-item multiple choice test comprised of items selected from the National Assessment Test for Science. Items on this test were drawn from those developed for 13 year-olds. The test was used in 8th grade science classes.

2. Science Process Inventory (Welch & Pella, 1967)

   This instrument consists of 135 agree-disagree statements that describe the assumptions, activities, products and ethics of science. The KR_{20} reliability is .86.

3. Science Attitude Inventory (Moore & Sutman, 1970)

   A Likert-type attitude measure designed to measure intellectual and emotional scientific attitudes. The 60-item test has a reported test-retest reliability of .93.

4. Learning Environment Inventory (Anderson, 1971)

   A 70-item measure of the students' perception of the social/emotional atmosphere in the class. Consists of 10-seven item scales labeled: Diversity, Formality, Friction, Goal-Direction, Favoritism, Difficulty, Democratic, Cliqueness, Satisfaction, and Disorganization.

5. Mathematics Inventory for Teachers

   The MIT is a 30-item instrument comprised in part of items from attitude items in the NLSMA Report (1968), and the International Study of Achievement in Mathematics (1967). The five Likert scales are designed to measure teachers': Attitude toward mathematics as a process, attitude toward classroom management practices, interest in teaching mathematics, rating of teaching practices, and rating of teacher concern for students.

6. Mathematics Attitude Inventory

   The MAI consists of six eight-item Likert scales and is designed to measure the secondary student's perception of: The math teacher, the value of mathematics in society, his enjoyment of mathematics, his anxiety toward mathematics, his self-concept in mathematics, and his motivation in mathematics. Pilot study reliabilities for the six scales range from .66 to .87 with four of the six having reliabilities greater than .80.

7. Teacher Questionnaire

   Two questionnaires were developed, one of 121 items for mathematics teachers and one of 137 items for science teachers. They were designed to obtain a broad view of the school's mathematics and science programs. The questionnaires focused on the teacher's perception of his personal background, skills and abilities, activities, and teaching environment.

8. Principal Questionnaires

Two 90-item questionnaires were developed, one for principals of mathematics teachers, and one for principals of science teachers. The questionnaires focused on the school and its respective science or mathematics department with an emphasis on programs and activities that might be carried out through the Comprehensive Projects. A number of items paralled teacher questionnaire items in order that teacher and principals perceptions could be compared.

9. Mathematics Achievement Test (NLSMA, 1968)

A. Form I (Grade 8)
This test is comprised of multiple choice items selected from a pool of eighth grade National Longitudinal Study of Mathematics Achievement (NLSMA) items. The selected items have correct response rates of between .50 and .90 and point biserial r's of at least 0.3.

B. Form II (Grade 11)
A 40-item, multiple choice test comprised of items selected from a pool of eleventh grade NLSMA items. The selection criteria were identical to those for the 8th grade test.

10. National Teachers Exam (Educational Testing Services, 1968)

A. Mathematics
The test consists of 120 multiple choice items, 10-15 per cent of which measure a teacher's knowledge of teaching secondary school mathematics. The remaining items provide a measure of the content of secondary school mathematics. The test has a KR_{20} reliability of .94, and offers normative data against which teachers of an impact region can be described.

B. Biology and General Science
A measure of knowledge about teaching biology and general science, and the content in biology and general science at the secondary school level. The test consists of 120 multiple choice items, has a KR_{20} reliability of .92 and offers normative data against which teachers of an impact region can be described.

C. Chemistry-Physics and General Science
This test measures knowledge of teaching chemistry, physics, and general science and the content of the three subjects at the secondary school level. The test consists of 120 multiple choice items, and has a KR_{20} reliability of .92.

11. Work Values Inventory (Super, 1970)

Part of the research effort of the project is focused on the values and attitudes of secondary school administrators. This instrument provides data on 15-scales related to work values. The reliability of the scales range from .74 to .88 with a median of .83.

Control Sample

Because of the need to examine changes on a given impact region during the four years of the Projects, appropriate control regions were identified. This process and its rationale are described in detail in Gullickson and Welch (1972). The control regions are basically the regions or states adjacent to the experimental regions. For example, the control region for Mississippi is Alabama; for San Jose it is Southern California, and so on. These regions were selected in order to reduce the variance due to regional differences across the country. The number of schools sampled within each control region is approximately one-half the number included in the experimental region. This decision was based on cost factors alone.

Figure 1 shows a map identifying the five Comprehensive Projects and their corresponding control regions.

Insert Figure 1 about here

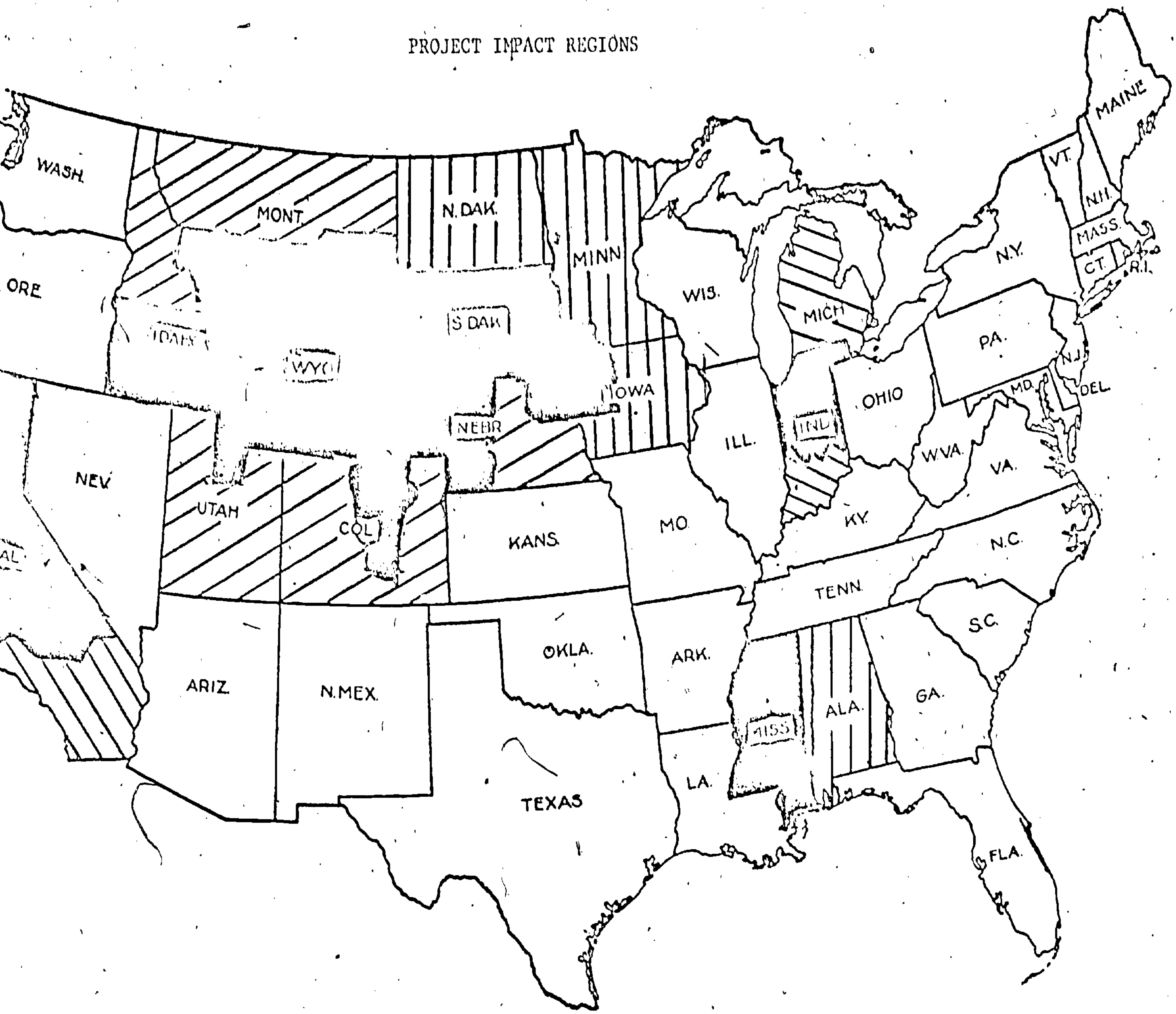
Testing Plan

Although the development and selection of the appropriate instruments represented major tasks, there was the additional problem of sampling the schools that would be involved in the study. The major considerations in this selection process are described elsewhere (Gullickson & Welch, 1972) and only will be summarized here. The primary consideration was the number required to detect at the $p < .05$ level a difference of one-half standard deviation. Table 2 presents the expected number of participants from each region and within each category. The states included in each experimental and control region are also listed.

Insert Table 2 about here

FIGURE 1

PROJECT IMPACT REGIONS



JECT REGION

NTROL REGION

NOTE: The Twin Cities were not included in the South Dakota control. Birmingham was not included in the Mississippi control region.

TABLE 2

Expected Sample Size

...tal Region	Principals	Teachers	Students	Control [illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
...r Calif.	1(0	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

...es an average class size of [illegible] for math and [illegible] for science.

Because of anticipated non-respondents, we over-sampled in each region by 40%. A key consideration in the sampling plan was to make certain that sufficient numbers of respondents were in each cell to provide valid needs assessment data to project directors.

Some idea of the extent of the project is revealed by considering that the total evaluation provides for testing nearly 800 principals, an equal number of teachers, and more than 21,000 students. In addition, data were gathered on 17 different instruments from schools scattered throughout 15 states.

Because of the twofold goal of this evaluation study, i.e., regional needs assessment data and pretest data, we decided to hold a series of regional meetings bringing together teachers, principals, project directors, and NSF representatives. Fourteen such meetings were held in the Spring of 1972. These meetings provided the opportunity for three activities. First, NSF education programs were explained in some detail to school personnel; secondly, Comprehensive Project Directors were provided with an opportunity to interact with school people in their respective impact regions; and thirdly, valuable test data were obtained from teachers and principals. In addition, we were able to provide teachers with packets of tests to administer to students in one of their math or science classes selected at random. Data were gathered from these classes using a system of random data collection within classes that increases the number of measures used, but minimizes testing time for any one student (Walberg & Welch, 1967). Data were collected simultaneously from randomly selected thirds of the class on three separate instruments. This system yields class means on three different tests, but requires only one 45 minute class period.

The set of regional meetings was coordinated by five regional coordinators. During these meetings, a total of 354 schools were represented. This is approximately 70% of the 515 we had hoped for. Although the number was somewhat disappointing, we were extremely pleased with the fine cooperation received from the parties concerned. We believe there is adequate representation from each region to provide useful needs assessment data to project directors.

Due to financial limitations, we decided early in the process to gather our control sample data through the mail. This meant foregoing achievement testing of teachers, but it did provide a relatively inexpensive way to gather the remaining data. As mentioned earlier (Table 2), 262 schools comprised the expected control group. Here also, we oversampled by 40% because of anticipated non-response. We received completed tests from 230 control schools, approximately 85% of the expected total. Again, it is our judgment that this number is adequate for comparison purposes.

Discussion

In this as in other evaluation strategies, the success of the plan is determined by the usefulness of the information to the several decision makers. The data gathered by this national design will be processed and analyzed during the next several months and presented to Comprehensive Project Directors and to the National Science Foundation.

We believe that answers must be found to several types of questions: Do principals hold negative attitudes toward science? What are the relationships between teachers and students attitudes toward mathematics? Is lack of subject matter competancy the primary problem in a given impact region, or is it low student interest? Our evaluation strategy is designed to gather data relevant to these and many other questions. If such information helps decision makers to achieve the goals of the Comprehensive Program, than this evaluation strategy will have been successful.

References

Anderson, Gary, The Assessment of Learning Environments: A Manual for the Learning Environment Inventory and the My Class Inventory. Halifax, Nova Scotia, Canada, February, 1971.

Gullickson, Arlen R. & Welch, Wayne W., Applying Experimental Designs to National Program Evaluation. Minneapolis, University of Minnesota, 1972.

Husen, T., International Study of Achievement in Mathematics, Vol. I. New York: Wiley, 1967.

National Assessment of Educational Progress. Report I, 1969-70, Science: National Results. Education Commission of the States, July, 1970.

Rutherford, F. James, Goals of Comprehensive Program. National Science Foundation, Washington, D. C., 1971 (Mimeo).

Walberg, Herbert J. & Welch, Wayne W., A New Use of Randomization in Experimental Curriculum Evaluation. The School Review, Winter, 1967, 75, No. 4.

Welch, Wayne W., Evaluation of the PSNS Course. Journal of Research in Science Teaching, 1972, 8.

Wilson, J.W.; Cahen, L.S.; & Begle, E.G. (Eds.) NLSMA Reports, No. 1, Part B. Stanford, California: School Mathematics Study Group, 1968.